Practical Assessment, Research & Evaluation

A peer-reviewed electronic journal.

Volume 14, Number 9, May 2009 ISSN 1531-7714

Gating items: Definition, significance, and need for further study

Wallace Judd, *Authentic Testing*

Over the past twenty years in performance testing a specific item type with distinguishing characteristics has arisen time and time again. It's been invented independently by dozens of test development teams. And yet this item type is not recognized in the research literature. This article is an invitation to investigate the item type, evaluate the contexts in which it may be appropriately used, and assess possible statistical and administrative ramifications of the item type. The nearest approximations to this item type in the literature are "non-compensatory items" or "conjunctive items", although for reasons that will become apparent it might be more appropriate to simply call them "gating items." Readers, researchers and practitioners are encouraged to address the issues of 1) How to evaluate and document the quality of these items for which traditional item statistics do not appear to be appropriate; and 2) How to incorporate gating items into the evaluation of the instrument.

A gating item is an item which, if failed, fails the examinee for the entire test. If passed, the examinee's score on the rest of the test will be evaluated. Passing the gating item does not assure passing the test; failing the gating item will fail the test. As will become apparent in the examples provided below, a gating item may occur at any time during the test, so it is not a filter through which an examinee must pass in order to take the rest of the test.

This paper presents examples of gating items, distinguishes them from other types of test questions, identifies some test and item analysis issues, and makes several recommendations for practice.

EXAMPLES

The first recorded instance I could find of a "gating item" is the Waterford fruit bowl (Waterford, 1996) used in the mid-14th century to determine whether an apprentice was ready to become a journeyman. If, within a week, an apprentice glasscutter could transform one of three glass bowl blanks into a fruit bowl, the glasscutter would become a master craftsman. Unfortunately, in this instance the single item was also the entire test, thus confounding item with test.

In the 20th century, one of the earliest and best-documented instances of a gating item occurs on the FAA pilot's flight test. During the flight test, the prospective pilot is asked to demonstrate proficiency in a number of flight procedures, including pre-flight inspection, takeoff, navigation, flight maneuvers, and stalls. At the conclusion of the test, the pilot must do one last thing – land the plane. If the pilot cannot land the plane in three tries, the FAA examiner takes over the controls, lands the plane, and fails the pilot – no matter what level of proficiency the pilot has exhibited in the prior exercises. Landing the plane is a gating item.

Foreign-trained veterinarians are given a practical exam by the American Board of Veterinary Medicine before they are certified to practice in the U.S. (E. Sabin, personal communication, March 7, 2007) The practicum includes seven stations at which examinees treat live animals exhibiting a variety of symptoms and requiring a variety of treatments. At one of the stations candidates are asked to spay a cat or dog. If the candidate puts the animal's life in jeopardy the attending veterinarian takes over and tries to save the animal, and the candidate discontinues practice and has failed the practicum. This is clearly a gating item.

In an exam of Linux system administrators, for example, candidates who can't add a new user to the system will fail the exam, no matter what other skills they exhibit. Obviously, no matter how well candidates can install printers, load balance or tune the system, or configure applications, all the benefits their skills deliver are useless if new users can't access the system (E. Liebovitch, personal communication, June, 2004).

Likewise, on an exam of Oracle system administrators, backing up the transactional database is a gating item. During the exam, a catastrophic failure of the operating system is induced. If the candidate hasn't backed up the system with each transaction, the system crash becomes an unrecoverable event. This is simply unacceptable for a competent database administrator. Again, no matter what other skills are exhibited, the candidate justifiably fails and cannot continue the exam, since any other skills are overwhelmed by the lack of a system backup (M. Serpe, personal communication, October, 2003).

In laproscopy, a candidate must be able to tie an inter-corporeal laproscopic suture. A candidate who is not able to tie the suture could never complete an operation, no matter how expertly performed. (R. Satava, personal communication, March 9, 2009)

The Red Hat RHCE exam, a candidate is presented with a system that has crashed. As frequently occurs in the real world, the prospective system administrator doesn't have a password for the system. If the candidate can't break into the system without a password, the candidate fails the exam, because no matter what skills of configuration, tuning, balancing or integration the candidate possesses, those skills are moot if he or she can't get into the system. (P. Childers, personal communication, May 19, 2005)

In the Chicago plumber's exam, the plumber must be able to fabricate a watertight system. The system is relatively simple and not contained within a domestic or commercial structure, but requires cutting pipe and soldering elbows, pipes and sleeves into a watertight system. If the plumber can't do that, he or she is just not a competent plumber. (R. Roberts, personal communication, March 10, 2009)

In a massage therapists practical exam, the massage therapist must ask the question, "Are you experiencing any pain in your body?" before beginning a massage. Without asking the question, the massage therapist may aggravate an existing injury by massaging the area over the injury. (W. Hogan, personal communication, February 7, 2008)

During the Landscape Architect practical exam, the candidate is asked to trim a bush with a chainsaw. If the candidate fails to don appropriate safety gear before using the chainsaw, the candidate fails the exam. (C. Chaffee, personal communication, September 29, 2008)

In the 2008 version of the NCCCO crane operator's exam, the candidate must put the headache ball in two 60-gallon oil drums separated by 180 degrees without knocking the drums over, in less than three minutes. The locations of the drums are carefully specified so that the crane operator must not only rotate the boom, but must simultaneously change the elevation of the boom. Failing to complete this task lost sufficient points that no amount of dexterity in the rest of the exam could lead to a passing score. (G. Brent, personal communication, June 21, 2007)

As you can see from the variety of these examples, gating items have arisen in a wide variety of certification and licensure settings. Gating items are a natural, often inevitable result of the conditions of performance.

As is evident from the references in the preceding paragraphs, not one of the examples above is cited in the literature. Consequently, for legal defensibility purposes, these examples don't exist.

GATING ITEMS, CRITERION REFERENCED ITEMS & PERFORMANCE ITEMS

Gating items, in all the instances of which this author is aware, are performance items. There is little opportunity for guessing because the response options are so extensive. Also, because the response modality is performance, the test developers assume that performance on a gating item is very likely to be indicative of performance subsequent to the test.

Gating items are always scored as domain-referenced items, not normed items, so hypothetically all candidates could pass them. It is the clear intent of the developers that the examinee exhibits a specific behavior.

Because of the critical ramifications of scoring a gating item, they typically have extremely objective scoring criteria. In the previously cited examples, probably the one requiring the most judgment to evaluate is whether the veterinarian candidate at the ABVM spaying station has jeopardized the animal's life. Landing a plane in three tries is not open to much interpretation. Nor is adding a new user to a Linux operating system. Acceptance of an intracorporeal suture could be a matter of judgment, but criteria such as knot size, tension on the knot ends before cutting, and holding power of the

suture can clarify the judgment. Whether the massage therapist asked the question about pain and whether the landscape architect donned full safety equipment are not open to interpretation. And whether the Red Hat administrator candidate can break into a system without the password is not ambiguous. Criteria for gating items are specifically designed to minimize the requirement for observer interpretation.

Different labels

Discussions with other psychometricians and practitioners have turned up several terms for these and similar items: domain critical items, non-compensatory items, critical items, mandatory items, and gating items.

The terms 'domain critical' or simply 'critical' items don't convey the absolute unmitigated requirement for passing that the term 'gating items' connote. As discussed by Friedman (1989), criticality is a matter of degree similar to importance or significance. As a matter of practice, criticality is often evaluated as a scalar in job task analysis.

The term 'mandatory' item calls to mind the 'mandatory' elements an Olympic skater must perform as part of the short program performed for the judges. Skating the elements is mandatory; scoring perfectly on them is not mandatory. Hence there is little parallel with these item types.

Readers familiar with medical and psychological testing literature may surmise that gating items are similar or identical to "critical items". There are subtle but important differences. While Newmark and McCord (1989), writing about the MMPI-2, state that "Critical items are frequently used as 'stop items' in screening patients," they also assert that "no empirical validation of these items has occurred." (p. 45) Further, they state that "Endorsement of any of the critical items should not be accepted as valid because an error or misunderstanding could have occurred" (p. 45). Finally, they state "In all cases, caution should be exercised when using critical items since single items are extremely unreliable indicators of psychopathology" (p. 45). Green (2000) in reviewing the MMPI-2 clearly describes critical items as being a part of a group of items with a cutpoint. Greene, in Appendix C (p. 572) identifies a variety of Critical Item Sets. Clearly, critical items identified for use in the MMPI-2 are not to be used singly to make a diagnosis.

The terms 'domain critical' or simply 'critical' items don't convey the absolute unmitigated requirement for passing that the term 'gating items' connote. As discussed by Friedman (1989) in the context of standard setting for health certification examination, criticality is a matter of degree similar to importance or significance, not as a binary trait (p. 4).

The term gating items seems particularly appropriate when reviewed in light of a logic gate: the item is an AND gate. The examinee must pass this item AND meet any other conditions required for passing. Parallel to a gating item, a logic gate can be placed at any place in the schematic and need not be positioned at any specific location in the circuit. Hence, 'gating items' seems an apt appellation.

LITERATURE REVIEW

It is difficult to find citations for gating items since they have not been documented by testing academicians.

Cizek and Bunch (2007) recognize nine different methods of standard setting for exams; gating items are not mentioned in any one of the methods. Baker and Kim's (2001) comprehensive treatment of Item Response Theory neglects to mention gating items. Shrock and Coscarelli (2005) do propose a two-tiered scoring system in which 'non-substitutable' skills are required to be performed with 100% accuracy (p. 189). The only other mention of gating functions appears to be in multistage testing where the results of one tier may gate access to another stage of testing.

Wainer, Bradlow, & Wang (2007), state that "The testlet can be as small as a single item (although in this extreme case, none of the advantages discussed here would hold), as large as the entire test, or anything in-between." (p. 57) Viewing a gating item as a degenerate (in the geometric or mathematical sense) testlet admits all the apparatus for evaluating a testlet, such as evaluating a passing score with a posterior probability of passing.

The appropriateness of compensatory and conjunctive scales has been discussed extensively (Way, Ansley, & Forsyth, 1988; Bolt & Venessa; 2003). Compensatory scales allow strength in one set of skills to make up for deficiencies in others. Conjunctive scales require demonstrated proficiency in one skill set that cannot be compensated for by proficiency in other areas. In discussing compensatory versus conjunctive models, Mehrens and Phillips observe (2008) "If there is a nonlinear relationship between one of the predictors and the criterion measure, it would be a violation of the model's assumption to use a linear regression method. … If the relationship is not at least monotonically

increasing, no compensatory model would be appropriate." (p.279) This provides a foundation for deciding whether a conjunctive model is appropriate, although they do not appear to consider the case of a single-item conjunctive test.

In discussing complex, innovative item types, Williamson, Bauer, Steinberg, Mislevy, & Behrens (2007) state, "A typical assessment uses not one but many task models, each capable of generating many tasks according to model specifications." (p. 6) While this may indeed be true for many task models, task models for gating items frequently can generate only a single task from the model specification.

Bolt and Lall (2003) state, "Because noncompensatory models often include component-specific difficulty parameters, their estimation requires sufficient variability in the relative difficulties of components across items to identify the dimensions." (p. 396) They go on to suggest Bayes factor analysis to evaluate model conformity to simulation data. (p. 407) This is a promising approach, but one that may prove intractable due to discontinuous ability distribution parameters.

While each of the methods cited above provides tantalizing hints of approaches that may work, none directly addresses the issue of a single item which can result in failure for the entire test.

There are a number of reasons gating items may not appear in the literature.

The first is that gating items are not widely used in educational settings, and much of the testing literature addresses issues critical to educational settings. Educational settings frequently cannot fund the equipment required to set up a performance test, nor can they assemble the experienced personnel required to judge the responses of a performance test. For most domains, performance tests cannot be cost-effectively scaled to the large numbers required of educational institutions.

Another possible reason gating items are not represented in the literature is that the people who created them are not psychometricians. They are plumbers, pilots, programmers, massage therapists, crane operators or even glasscutters. Consequently they are naïve about the theoretical issues their items raise, and unfamiliar with the venues in which to raise them. Moreover, they have no incentive to discuss them in a literary context.

Perhaps a third possible reason gating items may not be introduced into the domain of legitimate item types is that they are an affront to some practitioners' sensibilities. Some testing advocates steeped in the multiple choice environment feel it must be unfair to fail an examinee on the evidence of a single item. Their reactions to gating items are understandable, because in the multiple choice world one feels that all items can have sufficient compensatory evidence presented to overrule their indications. The evidence of a multiple-choice item is mitigated both negatively by the chance of guessing and positively by the possibility of inattentive or inadvertent failure. Either way, it seems reasonable to collect additional evidence before rendering a verdict on the item. However, performance items are different in that one may reasonably assume that what is demonstrated during the test is what the candidate will do in practice – and the performance evidenced during a failed gating item is so detrimental to either the candidate or his client that the aspiring practitioner should not be certified to practice.

A fourth and final reason may be that these items don't fit into the theoretical framework of IRT and adaptive testing. Clearly, if one could administer these items adaptively, one would administer the item at the beginning of the test. But because performance tests are sequenced by the nature of the task, items cannot be presented at the administrator's convenience. One can't ask a surgeon doing laproscopic surgery to suture a patient before making an incision. One can't ask a pilot to land the plane before taking off. One can't ask a database system administrator to make a backup before the system is installed. Consequently the timing of the administration is beyond the test developer's powers. And so gating items are off the table for people wishing to administer adaptive testlets, adaptive testing under either IRT or Rasch modeling, or even decision theoretical adaptive testing.

RATIONALE FOR USE

Why would one include gating items on a test? The testing literature doesn't afford defensibility. Typically, the P-Values for gating items are between 0.97 and 0.98, so they would not be selected for IRT information. And a P-Value of 0.97 predicts that the point-biserials will be near zero if the entire exam is given despite the score on the gating item. These seem like good psychometric reasons to exclude gating items from administration.

Gating items are often given because they provide information that cannot be derived from multiple choice items.

In *The Knowing-Doing Gap*, authors Pfeffer and Sutton (2000) explore numerous reasons corporate employees may not do what they know how to do. In other instances, knowledge of each of the components of a complex task does not necessarily mean that they can be assembled successfully into a complex performance. To review some of the examples above from this perspective, how many system administrators could not correctly answer the following question:

> How frequently should you back up a transactional database?
>
> A. Weekly B. Daily C. Hourly
> D. Each transaction

Yet candidates taking a major database certification program were reported to occasionally fail to actually back up the system. Laproscopic surgeons may be able to articulate the sequence of actions required to tie an intracorporeal laproscopic suture, but can not do it with laproscopic instruments. And a massage therapist may know, when asked, that she is required to inquire as to whether the patient is experiencing pain. But the massage therapist may fail to ask that simple question before beginning therapy.

Clearly, as evidence of competence a gating item provides information that cannot be obtained by a selected response item. And because of time, administrative constraints, or because the item is self-prompting, it may not be feasible to give the item more than once during a test.

ISSUES

What, then, are some of the issues which creators and users of gating items must confront? Below the issues are divided into item analysis, development, cutpoint and test-level issues.

Item Analysis Issues

How is one to compute a meaningful point-biserial coefficient? The results from the classical formula (Guilford, 1965):

$$r_{pbi} = \frac{M_p - M_t}{\sigma_t}\sqrt{pq} \quad \text{(Eq. 1)}$$

are zero when the test is terminated because of failure on an item, since in that case $M_p = M_t$.

One could score the test total score as zero, but this would dramatically penalize the point-biserial correlations for all other items, since the standard deviation of the test would rise substantially. One could also argue that all items up to and including the gating item should be scored, in which case the point-biserial is substantially different from zero. All three of these interpretations are open to question, and at this point there is no resolution about what statistics to report for gating items. One reasonable course of action would be to report only a passing percentage for gating items. This still leaves unresolved the question of whether the gating items in a test should contribute to its internal consistency reliability. Since internal consistency reliability statistics are based on inter-item correlations, and since the inter-item correlation for a gating item is low, including a single gating item in a relatively short performance test would substantially reduce the measured reliability of the test.

Gating Item Development

Two examples may serve to illustrate how gating items occur naturally in the development of a performance test. Cronbach (1970) discusses an example in which the Navy wanted to train sonar operators. A composite battery of tests was used for selection. When many men failed because of very poor tonal judgment, it was determined that their high mechanical comprehension scores raised their composite scores enough to conceal their tonal weakness. Cronbach states "Such men, despite an adequate 'average' ability, were doomed to fail in sound training whereas they would have been excellent in engineering, radar maintenance, or navigation. … Ultimately, a multiple-cutoff procedure was adopted." (Cronbach, p. 437) It would seem that a gating item involving tonal discrimination was called for under the circumstances, though Cronbach doesn't detail the resolution.

Another example occurred as the author was developing a performance test of Microsoft Word®, in a study comparing multiple-choice, simulation and performance tests. Neither the multiple-choice nor simulation tests contained gating items. But each scenario in the performance test required candidates to open a file, make appropriate edits, and save the resulting file. It quickly became apparent that if the candidate could not open or save a file, the candidate would fail the test. A macro could have automated the process, but further consideration made it clear that these were appropriate gating items. Indeed, if an examinee can not save edits, what good are the skills exhibited? Consequently, the

requirements to "open file" and "save as" became gating items.

Cutpoint Determination

Conventional standard setting procedures have no contingency for dealing with gating items.

Cizek and Bunch (2007) review nine standard setting methods, none of which explicitly or implicitly address the issue of gating items or critical items. The most widely used standard-setting method, the Angoff method (Angoff, 1971) assumes that scores on all items are compensatory and thus does not work with gating items. Friedman (1989) discusses a procedure for incorporating 'Critical items' as part of a standard setting process for medical certification. But Friedman conceives of critical items as being used as part of a set of critical items with a separate cutpoint established for the set, as opposed to binary pass-fail indicators. In the *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association & National Council on Measurement in Education, 1999) the single mention in the chapter, "Testing in Employment and Credentialing", is "When evidence of validity based on test content is presented for a job or class of jobs, the evidence should include a description of the major job characteristics that a test is meant to sample, including the relative frequency, importance, or criticality" (p. 160). Shrock and Coscarelli (2005) use the term "non-substitutable skills" for gating items and suggest that "you can partial out the non-substitutable skills and establish a two-tiered scoring system in which a score of 100% is required on the non-substitutable items, and a given percentage is required for the remaining items." (p. 189)

Test-Level Issues

How does one go about documenting that a gating item is a good item? What is adequate evidence of the criticality of content that should be persuasive to a jury of peers or to a court of law? A reasonable method would be to assemble a group of content experts, and have them vote on whether or not the item should be a gating item. The group needs to achieve more than consensus on the issue; they should reach a unanimous decision that the item is indeed of such importance that it is a gating item. If the decision is not unanimous, then it may be reasonable to instead define a high weight in a compensatory scoring rubric.

How would one evaluate equivalent forms? If a gating item is so critical to evaluating professional practice, is it reasonable to create an equivalent form of the item?

How is one to account for the test scores of examinees who are allowed to continue after failing a gating item? Are these legitimate scores that should be included in the mean score? Or are they only valid as pass/fail scores? How do you report to a candidate that a test score above the cutpoint resulted in failure on the exam? What if an examinee has taken a portion of the exam prior to encountering and failing a gating item? Is the examinee's test score the score for the portion of the exam the candidate was allowed to take?

What is the proper procedure for computing Cronbach's alpha when an exam includes a gating item? If the exam was terminated on failure of the item, clearly the correlation of the gating item with all subsequent items is undefined. If the exam is continued after the gating item is failed, computation of alpha is possible but meaningless, since passing or failing subsequent items is meaningless.

These issues present serious unresolved issues for psychometricians, test developers and researchers to deal with.

Fortunately, in actual practice gating items are infrequent, typically comprising a small percent of the items on a test form. And, indeed, they are so fundamental to the practice being evaluated that they typically have a low observed frequency of failure. Nonetheless, they are a legitimate component of performance testing, and need to be incorporated within the theoretical framework of standard setting, exam evaluation and item evaluation.

SOME IMPLICATIONS FOR PRACTICE

Despite the unresolved theoretical issues discussed above, gating items will continue to appear in performance exams. For those developing performance exams, following are a few suggestions:

1. Recognize gating items that are intrinsic to the content of the exam. Conversely, do not try to find them if they are not an inevitable part of the domain being tested. Gating items should arise from a dire need for patient or client safety. When catastrophic results, such as a patient dying or a crane collapsing, are the result of a single action, that action may reasonably be the content of a gating item.

2. Do not eliminate gating items because their P-Values are high. A P-Value of 0.95 to 0.98 is quite normal for a gating item. The significance of a gating item cannot be summarized in statistical terms. The rationale for including a gating item arises from its content and consequences, not from the mathematical relationship of the item to other items or total test score.
3. Include the gating item as an explicit part of the scoring rubric. One could conduct an Angoff evaluation of exam items excluding the gating items, then state a scoring rule in the form: A passing score consists of passing all gating items and achieving a 78% score on the remaining items. The scoring rubric should also document that the expert committee unanimously endorsed the content and criticality of the gating item.
4. Be sure that the directions for the gating item are absolutely clear. The results of these items must be unambiguous because they have the most severe consequences of any items on the exam. It would be unacceptable to fail an examinee on a gating item simply because the candidate did not understand the item instructions.
5. Similarly, the criteria for scoring a gating item must be unambiguous. For example, the statement, "The candidate must wear appropriate safety gear before using the chainsaw." is not adequate because of the ambiguity of 'appropriate'. An improved form would be, "The candidate must be wearing a helmet, goggles, gloves, safety boots and chaps before starting the chainsaw."
6. It is possible to abuse the use of gating items. For test developers to develop poor gating items that would be a major disservice to test takers. Gating items must express a consensus of professional practice and not be the opinion of just one segment of the profession, no matter how influential.

REFERENCES

American Educational Research Association, American Psychological Association, National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: Authors.

Angoff, W. H. (1971). Scales, norms, and equivalent scores. In R. L. Thorndike (Ed.), *Educational measurement*. Washington, DC: American Council on Education.

Baker, F. B. & Kim, S. (2004). *Item response theory; Parameter estimation techniques, Second Edition*, Marcel Dekker.

Berk, R. A. (1984). Selecting the index of reliability. In R. A. Berk (Ed.), *A guide to criterion-referenced test construction*. Baltimore: Johns Hopkins University Press.

Bolt, D. M. & Venessa, F. L.; (2003). Estimation of compensatory and noncompensatory multidimensional item response models using Markov chain Monte Carlo, *Applied Psychological Measurement, 27,* 395-414.

Cizek, G. J. & Bunch, M. B. (2007). *Standard setting, a guide to establishing and evaluating performance standards on tests*; SAGE Publications, 2007.

Cronbach, L. J. (1970). Essentials of psychological testing, Third Edition. New York: Harper & Row

Friedman, C. (1989). Critical items and standard setting for health certification examinations. (1989). San Francisco: Paper presented at the Annual Meeting of the National Council on Measurement in Education.

Greene, R.L. (2000) *The MMPI-2: An interpretive manual, Second Edition*. Boston: Allyn and Bacon.

Guilford, J. P. (1965). *Fundamental statistics in psychology and education*, *Fourth Edition*. New York: McGraw-Hill.

Liebovitch, E. (2004). Personal communication.

Mehrens, W.A. & Phillips, S.E. (1989). Using college GPA and test scores in teacher licensure decisions: Conjunctive vs. compensatory models. *Applied Measurement in Education, 2*, 277-288.

Newmark, C.S. & McCord, D. M. (2000). The Minnesota Multiphasic Personality Inventory-2 (MMPI-2). In C. S. Newmark (Ed.), *Major Psychological Assessment Instruments, Second Edition*. Boston: Allyn and Bacon.

Pfeffer, J. & Sutton, R. I. (2000). *The knowing-foing hap: How smart companies turn knowledge into action.* Boston: Harvard Business School Press.

Shrock, S. & Coscarelli, W. (2005). Criterion referenced test development. Washington, DC: International Society for Performance Improvement.

Wainer, H., Bradlow, E. T., & Wang, X. (2007). *Testlet response theory and its applications*, Cambridge: Cambridge University Press.

Way, Walter D., Ansley, T N., & Forsyth, R. A. (1988). The comparative effects of compensatory and noncompensatory two-dimensional data on unidimensional IRT estimate. *Applied Psychological Measurement*, *12*, 239-252.

Williamson, D. M., Bauer M., Steinberg, L. S., Mislevy, R.J., & Behrens, J. T. (2007). Design rationale for a complex performance assessment, Draft to appear in *The International Journal of Testing.*

Waterford (1996). Promotional video obtained in Waterford Crystal, in Waterford, Ireland. Waterford uses the fruit bowl for a final exam, because it requires 49 separate cuts. However, at Kinsale the fruit bowl separates an apprentice from a journeyman; a Master bowl separates a journeyman from a master craftsman. "A bowl is usually the item favoured to test an apprentice glass cutter. This test is given after serving five years. Then after ten years, a more difficult bowl is used to test for a Master Cutter. This has become known as the Master Bowl." http://www.kinsalecrystal.ie/bowls/bowls.htm

Citation

Judd, Wallace (2009). Gating items: Definition, significance, and need for further study. *Practical Assessment, Research & Evaluation*, 14(9). Available online: http://pareonline.net/getvn.asp?v=14&n=9.

Author

Wallace Judd
Authentic Testing Corporation
E-mail: WJudd [at] AuthenticTesting.com